



Part 1: Introduction, Guide to the Documentation and Validation Statement



February 14, 1991

T. Hull

3-60-91-12

DATA FROM WAR RELOCATION AUTHORITY FORM 26:
EVACUEE SUMMARY DATA ("LOCATOR INDEX")

Guide to Documentation

This guide to the WRA Form 26 Evacuee Summary Data dataset is intended to provide pertinent information regarding the use of the documentation package. Elements of this documentation package were collected from a variety of sources including the Bancroft Library at the University of California at Berkeley and the Civil Reference Branch of the National Archives (NNRC). In addition, addendum to this documentation package are the result of a concerted effort by staff in the Center for Electronic Records, Archival Services Branch (NNXA). The documentation collection is comprised of the following:

- The introduction to the documentation outlines some of the larger scale problems encountered during verification of the dataset. Associated with this piece of documentation is the verification statement, describing in some detail the particular discrepancies encountered during hand validation between the sample dump of the dataset, the codebook, and the punch cards.
- The coding and keypunch instructions are copies of the original documents provided to employees of the WRA responsible for coding the original WRA Form 26 and keypunching the data onto IBM cards. These documents are provided to give some background regarding the WRA project.
- The main codebook is that used for verification of the dataset. It provides the user with a guide to translating the codes used by the WRA for keypunching the data. The verification statement should be consulted to provide the user with additional information about some discrepancies which arose between the codebook and dataset.
- Part four of the documentation provides a translation table to the geographic codes used in columns 22 through 26 of the dataset.

- Part five of the documentation package provides a translation table to the occupation codes used in columns 55-57, 58-60, 61-63, 64-66, and 67-69 of the dataset. In addition, original coding instructions for these items are included.

- The last section of the documentation package includes some additional information regarding the information in columns 70 through 75 of the dataset (the file number) and the individual number (columns 39 through 44). In addition, this section includes a table identifying sects of the Buddhist religion and the 'age from emperor at year of birth' translation table.



Washington, DC 20408

March 14, 1991

T. Hull

3-60-91-012

DATA FROM WAR RELOCATION AUTHORITY FORM 26: EVACUEE SUMMARY DATA ("LOCATOR INDEX")

Introduction

The Evacuee Summary Data dataset contains information about Japanese-Americans interned in 10 camps during World War II by the War Relocation Authority (WRA). The original information was collected on WRA Form 26, coded by employees of the WRA, and keypunched by the agency. Apparently, two sets of cards were created: one set is in the custody of the National Archives, and the second was deposited at the Bancroft Library of the University of California at Berkeley following the termination of the WRA. (See WRA Preliminary Inventory #77, Entry #23 (1955) and Japanese American Evacuation and Resettlement: Catalog of Material in the General Library (1958, University of California)). The cards in the possession of the National Archives are apparently the original set produced by the WRA, based on the opinion of Dr. John Modell of Carnegie Mellon University, Department of History. In addition, the National Archives punch cards were manually annotated by the WRA with information about internees aliases, date of departure and place of destination, and other uncoded information. According to Dr. Modell, the Berkeley cards did not contain such additional information and were a copy of the National Archives cards.

The second set of cards were "read" to tape at U.C. Berkeley during the 1960's and a copy of that tape is currently in the custody of the National Archives. Following the decision to distribute reparations to former internees during the 1980's, the Civil Rights Division, Department of Justice acquired a copy of the dataset from U.C. Berkeley. The Justice Department subsequently transferred a copy to the National Archives. The National Archives set of punch cards were apparently never "read" to tape. Therefore, persons wishing to compare the National Archives punch cards to the electronic dataset should be aware of potential inconsistencies and/or gaps in the information. A number of these are noted below as they were discovered during validation.

The following sections outline some of the 'macro' issues that should be made known about the dataset. These issues were discovered while staff at the Center for Electronic Records

undertook validation processes on the file. More detailed issues regarding inconsistencies in coding and a translation table for actual codes appearing in the dataset appear in a separate document, the validation statement. The remainder of this document discusses those issues regarding the National Archives punch card deck and dataset as a whole, and how the two items relate or do not relate to one another. Among them are issues concerning extraneous records or information in the digital file and information contained in the original National Archives punch cards not included in the digital file.

1) Extraneous Records or Information

Users should be aware of a number of inconsistencies discovered during the preliminary evaluation. The first ten records of the file are either blank (no's 1, 3-8) or have information that is unintelligible (no's 2, 9-10). In addition, records 36-41 contain information that is apparently part of a tape copying routine used at some point in the history of the file and accidentally inserted into the file. These records should be disregarded.

The dataset is presented as a sorted file, sorted ONLY by the last name of the internee. Note that on the sample dumps of the first and last 5 blocks of the dataset (250 records each), the records 11 through 35 at the beginning of the file have a blank () or ampersand (&) in the first column. The file then continues with records in alphabetic order. The last 28 records of the file have numerals in the first column. As part of the validation process, a few of the anomolous records at the beginning and end of the file were checked for in the National Archives set of punch cards. Curiously, none were located.

2) Excluded Records or Information

A lot of what was learned about the electronic version of the WRA file came from comparing the sample dump to the set of punch cards in the National Archives. This provided the opportunity to undertake an evaluation of the extent of the information apparently missing from the electronic dataset, as derived from the Berkeley set of punch cards, but included in the Naional Archives punch card decks.

During the validation process, approximately fifty records from the sample dump were compared in detail to the original National Archives punch card record. As a result of this comparison, it was noted that manual or handwritten corrections to the typed information on the punch card (i.e. year of birth or a mis-spelling of the internee's name) were not corrected. That is, no new card was punched to replace erroneous information about the internee. Therefore, the electronic dataset may contain erroneous information which was corrected only through annotations to the cards.

A second difference between the National Archives punch cards and the digital version is that the punch card decks contain what may be termed 'cross-reference' cards for additional information about internees; particularly aliases. Associated with this issue is the inclusion on the cards of handwritten information about the internee which was not coded. This includes such items as when the internee left the camp, what the intended place of destination was, and information about transfers between camps. This sort of information is not a part of the digital database.

Finally, note that the file does not include information for those individuals who were born in the internment camps. Those cases are recorded in the National Archives punch card set, by use of a card with a blue strip across the top, but that information was apparently never keypunched and subsequently not entered into the present dataset.

3) Conclusions

Users of the Data from WRA Form 26 dataset should be aware of some of the problems that may be encountered. Firstly, due to limited verification procedures, the National Archives did not determine the completeness of the dataset. In addition, the file may contain erroneous information about individuals, including entirely blank or unintelligible records. A researcher should be prepared to deal with these problems during preliminary data analysis.

Persons interested in using the information in the electronic dataset for geneological purposes should be aware of sources available in other divisions of the National Archives or Bancroft Library and should consult those to verify the information. In particular, it should be noted that for researchers interested in information about those born in the camps, this file does not include those records.



February 14, 1991

T. Hull

3-60-91-12

DATA FROM WAR RELOCATION AUTHORITY FORM 26:
EVACUEE SUMMARY DATA ("LOCATOR INDEX")

Validation Statement

Total Number of Datasets: 1

Title: Data from WRA Form 26: Evacuee Summary Data
("Locator Index")

Number of Datasets: 1

Logical Record Length: 80

When the National Archives acquired custody of this file, the accessioning procedures called for a manual comparison of the documentation to a printout of a portion of the records in each data set. This manual comparison is referred to as a "preliminary assessment" or "validation." In this case, the comparison involved two distinct steps: 1) comparing the sample dump to the available codebook and 2) comparing the sample dump to the set of punch cards in the possession of the National Archives, on which the original coding was interpreted. This two step analysis helped to clarify some questions regarding differences between the sample dump and codebook. This detailed comparison involved approximately fifty records. This is a statement of the results of the preliminary assessment or validation.

The following are discrepancies between the documentation, National Archives punch cards, and sample dump of the data noted during hand validation:

1) The codebook outlines the use of a "x" and/or "u" as valid codes in columns 21, 22-26, 28, 29, 30, 31, 33-34, 35, 36, 37, 38, 45, 47-48, 49-50, 51, 52, 53, and 54, generally to indicate missing or unknown information. These special characters were to be punched in rows 11 and 12, respectively. The National Archives punch cards indicate a "\" or "*" in similar columns as translated across the top of the cards. The sample dump of the dataset does not contain such characters. Rather, a "-" or "&" is shown. Based on cursory analysis, these special characters are apparently used most frequently in columns 28, 29, 51, and 54. The following translation table may be used to understand the contemporary meaning of the original codes:

Table 1: Translation Table for Special Characters

Coding Character	Row Punch	Punch Card Character	Data Set Character
π	11	\	-
α	12	*	&

2) The variable appearing in columns 33-34, Year of First Arrival in Territorial U.S. for Foreign Born, uses three sets of special characters. For "π" and "α", use the information provided in table 1 for translation. In the case of the third possibility for this variable, where "--" is used to indicate those internees who were born in the U.S., but had visited Japan, no punches were made and subsequently the field is blank in the electronic dataset.

3) The codebook only outlines codes up through column 54. The printout indicates data were entered through column 80. Although a number of items were not clearly identified in the codebook, the meaning of their values could nevertheless be determined.

- The photocopy of a blank WRA Form 26 punch card shows that columns 55-57, 58-60, and 61-63 were to be used for coding occupations for which the internee was qualified and columns 64-66 and 67-69 were reserved for occupation codes for which the internee had potential. A check of the codes against the occupation code listing (Part 5) included with the codebook shows that the information reported is reasonable.

- In the case of the occupation codes, the appendix (Part 5) lists 'general' occupation codes with an "X" in the second position of the three digit variable. This was to be coded as a "π", thereby showing as a "\" on the punch card, and ending up as a "-" in the dataset. For example, for those internees who expressed an aptitude toward technical work, but had no particular experience, they may have been categorized by the interviewer as 0X7, coded as a 0π7, punched as 0\7, and now appear in the digital dataset as 0-7.

- Columns 70 through 75 is the six digit 'file number' variable; a unique identifier for each internee. Columns 76 through 80 is an additional identifier for which no information is known. This last item was not translated on the punch card set in the possession of the National Archives, but must have been punched on the set at U.C. Berkeley. According to Dr. Modell, this last data element was neither entered by his project, nor did he recollect the interpretation of it.